Medical Informatics in a United and Healthy Europe
K.-P. Adlassnig et al. (Eds.)
IOS Press, 2009

doi:10.3233/978-1-60750-044-5-215

An XML Schema for Automated Data Integration in a Multi-Source Information System Dedicated to End-Stage Renal Disease

Eric DUFOUR [a], Mohamed BEN SAÏD [b,c,1], Jean Philippe JAIS [b,c], Loic Le MIGNOT [b,c], Jean-Baptiste RICHARD [b,c], Paul LANDAIS [b,c]
[a] *SELIC, Janville, France*
[b] *Department of Biostatistics and Medical Informatics, APHP - Necker Enfants Malades Hospital, Paris, France*
[c] *UPRES EA 4067 - Paris Descartes University, Paris, France*

Abstract. Data exchange and interoperability between clinical information systems represent a crucial issue in the context of patient record data collection. An XML representation schema adapted to end-stage renal disease (ESRD) patients was developed and successfully tested against patient data in the dedicated Multi-Source Information System (MSIS) active file (more than 16,000 patient records). The ESRD-XML-Schema is organized into Schema subsets respecting the coherence of the clinical information and enriched with coherent data types. Tests are realized against XML-data files generated in conformity with the ESRD-XML Schema. Manual tests allowed the XML schema validation of the data format and content. Programmatic tests allowed the design of generic XML parsing routines, a portable object data model representation and the implementation of automatic data-exchange flows with the MSIS database system. The ESRD-XML-Schema represents a valid framework for data exchange and supports interoperability. Its modular design offers opportunity to simplify physicians' multiple tasks in order to privilege their clinical work.

Keywords. end-stage renal disease, multi source information system, interoperability, XML schema, electronic data exchange, internet, dynamic web server

1. Introduction

Redundancy of information is deleterious when it requires multiplying data entry on several information systems. Interoperability and data sharing between information systems in clinical settings are a necessity since they alleviate professionals' workload to allow further focus on patient data and decision making. The eXtensible Mark-up Language (XML) representation [1] brings an answer to data exchange between heterogeneous systems. It delivers key advantages in interoperability due to its flexibility, expressiveness, and platform-neutrality [2]. It offers perspectives for fine-

[1] Corresponding Author: Dr. M. Ben Said, Service de Biostatistique et d'Informatique Médicale, Hôpital Necker Enfants Malades, 149 rue de Sèvres, 75015 Paris, France; E-mail: bensaid@necker.fr.

grained disease and context representation independently from technical considerations of modelling and implementation.

An XML schema [3] was designed and tested to represent a standardized minimal medical record of patients with End Stage Renal Disease (ESRD) developed in the context of a national program called: Renal and Epidemiology Information Network (REIN) [4, 5]. A dedicated information system called the Multi-Source Information System (MSIS) [6] supported the REIN program.

This paper aims at describing the design of the ESRD-XML Schema, the processing of instantiated XML files to access to patient data and the methods used to store relevant information to update the MSIS production database.

2. Background

The REIN-ESRD patient record was elaborated by nephrologists and health care professionals. It was implemented in the MSIS which is in production since 2002 and deployed in 8 regions (over 23) in France. As of December 12th 2008, 34,378 patient records were registered in the MSIS production database including 16,610 records in the active file. Authorized users (584), including nephrologists (469), nurses (32), secretaries (73) and clinical research assistants (10) update MSIS ESRD-patient records.

Each patient record was organized in patient identification section and patient medical information section. This latter was organized around three major events which characterize the ESRD care process:

- Patient medical history, aetiology of ESRD and comorbidity at start of replacement therapy (dialysis or kidney graft).
- Recent medical observation with information about comorbidity and handicap evolution, facilities to access to the dialysis care unit and/or to the national waiting list of kidney transplantation.
- Actual renal dialysis method and context of treatment.

Admission, discharge and transfer event information are documented and annual follow-up observation are added periodically to the ESRD patient record.

3. Methods

XML is a standard for data exchange. The use of XML schema was privileged for its flexibility in representing specific knowledge domain. It provides means for defining the structure, content and semantics of XML documents [7]. It consists of components such as type definition and element declarations used to assess the validity of well formed element and attribute information items and specify augmentations to these items and their descendants.

In conformity with the REIN-ESRD-XML schema representation, a trusted legacy information system will periodically “send” instantiated XML messages with patient data extracted from local information system. If valid, patient data are automatically integrated in the MSIS production database. An ESRD-XML message carries information about only one patient. It has a fixed part made of a header, patient identification, care unit subsets and an iterative part made of description of an “ESRD-event”. Each event requires an action of creation, update or deletion of a patient record.

3.1. ESRD-XML Schema Design

ESRD-XML Schema is organized into a top schema document called ESRD_Events.xsd and subsets of XML schemas files called as "include". They correspond to a coherent set of clinical ESRD event information. Four XML Schemas files (header.xsd, attribute.xsd, generic.xsd and simpleType.xsd) correspond to a coherent subset of XML components, referenced by their names and relations.

- *ESRD_Events.xsd:* corresponds to the top ESRD-XML-Schema representation. It is represented in a typeMessage component, organized into header and one or more ESRD event components. Each event component declares the professional responsible for the patient information and the care unit where the ESRD event is observed. An ESRD event may declare all or part of the followings XML schema files: initialFile.xsd, followUp.xsd, transfer.xsd, decease.xsd, changeIdentity.xsd.
- *Header.xsd*: concerns the message envelope: it is of typeHeader and declares 4 components: message identifier, production time, and transmitter / receiver information;
- *Patient.xsd*: focuses on patient identification as an individual with demographic and patronymic information and as an ESRD patient;
- *InitialFile.xsd*: represents patient medical data at the initiation of ESRD replacement therapy;
- *FollowUp.xsd*: represents a dated medical observation event. It carries handicap and comorbidities information beside biology and therapy context. Semantics, possible values and ranges are declared for each entry;
- *Comorbidity.xsd* declares the definitions of 21 comorbidities' concepts;
- *Transfer.xsd:* the content allows indicating the new dialysis unit taking into account the positioned date of transfer;
- *Decease.xsd:* indicates the context and the medical causes in case of decease;
- *ChangeIdentity.xsd*: allows correcting patient identification data such as name, surname and place of birth.
- *Attribute.xsd, genericType.xsd, simpleType.xsd:* these files of include type allow homogenizing the iterative types of data used for the definition of attributes, complexTypes or simpleTypes.

3.2. Dictionaries

We used standard thesauri, maintained by institutional and academic organizations: the International Classification of Disease (10^{th} version), the French Thesaurus of Nephrology, the definition of the French districts according to the list of the National Institute for Statistics and economical studies (INSEE).

3.3. Tests

We proceeded in two phases to test the ESRD-XML-Schema model: firstly, we instantiated ESRD-XML documents with anonymized patient data. Secondly, we simulated the process the other way; instantiated XML-documents are generated by a local application carried ESRD-patient data to integrate in the MSIS database.

Parsing instantiated XML documents confirmed the correct format of the XML file and split it into its constituent elements. We used the Java library Document Object Model (DOM) parser to read the XML document in a tree structure [8].

A DocumentBuilder object obtained from a DocumentBuilderFactory allowed reading an XML document. The Document object is in-memory representation of the tree structure of the XML document. It is composed of objects, the classes of which implement the Node interface and its various sub-interfaces. The DocumentBuilder object first verifies the coherence with the ESRD-XML schema as referenced in the link of the XML document header. It eventually submits unconformity errors. The parser browses the tree structure representation in multiple manners in order to extract the child nodes Elements and attributes contents and values. It identifies iterative information subset contents and related instructions. A first parse looked for "ESRD_Event" node elements and kept track of their number and their "action" attributes. Alternatively, knowing the names of elements and attributes, we accessed directly to their content as appropriate.

Object model data representation was adopted to store information extracted through the parsing process. Information of the same type such as: Patient, CareUnit, or ESRD_comorbidity, were organized into object classes. The advantage was to work with fixed-length record format to store and process data of the same type. Every object [9] has a class, that defines its data or field variables (private String birthName;), and behaviour defined in the methods (public String getBirthName(){}, public void setBirthName(){}, etc).

A final step consisted of mapping the objects instance variables to the MSIS data model and making necessary controls before updating the MSIS database.

4. Results

All the parsing tests of XML instances were validated. The instantiated xml file corresponding to the message ESRD_Events.xsd issued the most complete patient record represented in conformity with the ESRD-XML schema. The successive tests of instances corresponding to the initialFile, followUp, transfer, decease and to the changeIdentity schemas carrying real ESRD patient data, were validated as well. ESRD-patient data extracted from automated XML document parsing were stored into data class objects and mapped to MSIS data model. It successfully served to update the production database and enrich the ESRD-patient cohort. Integration of the ESRD-XML documents processing into the MSIS application system and ESRD-XML Schema publication are in process.

5. Discussion and Conclusion

The direct involvement of the nephrologists in using MSIS to update patient information varies according to their workload. Feeding different types of information systems impairs medical activities. XML offers a solution to exchange information on the basis of a share schema. Thus, information already stored in a database can be transferred to another without repeating data entries.

XML schema representation helped eliciting the model in an abstracted way, easy to read and simple to exchanging information. It provides means of verifying the validity and the coherence of XML-documents with patient data content. XML might have a key role in interoperability given its flexibility, expressiveness, and platform-neutrality. It offers interesting perspectives for exchanging information in a country or for international survey purposes. We extended the use of XML schema exchange model to other medical domains, currently genetic rare diseases. ESRD-XML schema supports both knowledge and application domains. It provides basis to enrich the health care communities' efforts in research and innovation such as in Semantic Web for Health Care and Life Sciences Interest Group.

Acknowledgments. The nephrologists in charge of the ESRD units in Limousin, Languedoc-Roussillon, Champagne Ardenne, Centre, Provence-Alpes-Côte d'Azur, Basse-Normandie, Midi-Pyrénées, Ile-de-France are warmly acknowledged for their fruitful cooperation and comments as well as JP Necker and X Ferreira. This work was supported by a grant of Paris Descartes University.

References

[1] http://www.w3.org/TR/xmlschema-0.
[2] Matsa, M., Perkins, E., Heifets, A. et al. (2007) A High-Performance Interpretive Approach to Schema Directed Parsing. In Williamson, C.L. et al. (Eds.) *Proceedings of the 16th International World Wide Web Conference WWW 2007,* ACM, Banff, 1093–1102.
[3] http://www.w3.org/TR/xmlschema-1#d0e504.
[4] Landais, P., Simonet, A., Guillon, D., Jacquelinet, C., Ben Saïd, M., Mugnier, C., Simonet, M. (2002) SIMS@REIN: Un système d'information multi-sources pour l'insuffisance rénale terminale. *Comptes Rendus Biologies* 325(4):515–518.
[5] http://www.soc-nephrologie.org/PDF/enephro/registres/guide_REIN/2009-guide.pdf.
[6] Ben Saïd, M., Simonet, A., Guillon, D. et al. (2003) A dynamic Web application within an n-tier architecture: A multi-source information system for end-stage renal disease. *Studies in Health Technology and Informatics* 95:95–100.
[7] Martins, W., Neven, F., Schwentick, T., Jan Bex, G. (2006) Expressiveness and complexity of XML schema. *ACM Transactions on Database Systems* 31:770–813.
[8] Hortsmann, C.S., Cornell, G. (2005) *Core Java 2 – Volume II – Advanced Features*, Sun MicroSystem Press, Prentice Hall Title, 879–934.
[9] Hortsmann, C.S., Cornell, G. (2005) *Core Java 2 – Volume I – Fundamentals*, Seventh edition, Sun MicroSystem Press, Prentice Hall Title, 93–149 & 619–706.